Capstone Final Report
Factors Influencing Student Success

Yenmin Young

Introduction:

Background

Education plays a pivotal role in shaping a young person's future—academically, socially, and professionally. Graduating high school, enrolling in higher education, or securing stable employment are all widely accepted indicators of student success. Yet, the pathway to these outcomes is rarely linear and never universal. Developing critical thinking skills, gaining knowledge, and cultivating a strong work ethic are essential parts of the educational journey, but how students reach these milestones often depends on a complex web of influences. Understanding the factors that impact student performance is key to ensuring that all learners are equipped with the tools and opportunities they need to thrive.

A student's learning experience is profoundly shaped by both internal and external influences. Family background plays a major role—whether parents value education and are actively involved in their child's development can have significant effects on motivation, access to resources, and emotional support. Personal goals also matter: Does the student see education as an investment in their future? Are they intrinsically motivated to learn, or simply going through the motions due to obligation or external pressure? Financial barriers can further complicate this landscape, limiting access to tutoring, extracurricular experiences, healthier living conditions, and even basic necessities like transportation or stable housing. The way students allocate their time—balancing school, work, and social obligations—can greatly affect their academic engagement and outcomes.

These challenges do not occur in isolation, and their impact ripples out across various stakeholders. Students are at the center, of course—their futures depend on whether or not they can overcome these obstacles. Families, especially parents, have a vested interest in understanding how to support their children more effectively. Educators and administrators must ask how they can structure learning environments that recognize these factors and proactively respond to students in crisis. For policymakers and educational services, the question becomes: How can systems of support—financial, emotional, academic—be more equitably distributed to close the gap between potential and performance?

In Portugal, these questions are particularly relevant, given its diverse population and unique educational landscape. By examining key influences such as demographics, academic habits,

and social and familial structures, we can begin to identify which students are most at risk of failing. This understanding lays the groundwork for targeted interventions: proactive systems that not only identify struggling students early on, but also provide them with actionable steps and resources to improve their chances for success. Through this lens, we shift from simply observing outcomes to actively shaping them.

Guiding Questions

- What factors have the largest influence on a student's success?
- Which students will pass based on their information, background, and lifestyle?

Data:

Datasets

This project utilizes two datasets from Portuguese secondary schools—one focusing on Math and the other on Portuguese language classes. Each dataset includes student demographic data (age, gender, parental education and employment), lifestyle indicators (alcohol consumption, study time, social activities), and academic history (grades across the semester, prior failures, absences). The final grade ('G3') determines whether a student passed (score ≥10) or failed (<10).

Data Wrangling

The datasets were merged to unify students across both subjects. Cleaning was minimal—there were no duplicates or missing values, and all datatypes were appropriate. A new binary column was created to indicate whether a student passed or failed each course.

Exploratory Data Analysis

An initial analysis of grade distributions revealed that:

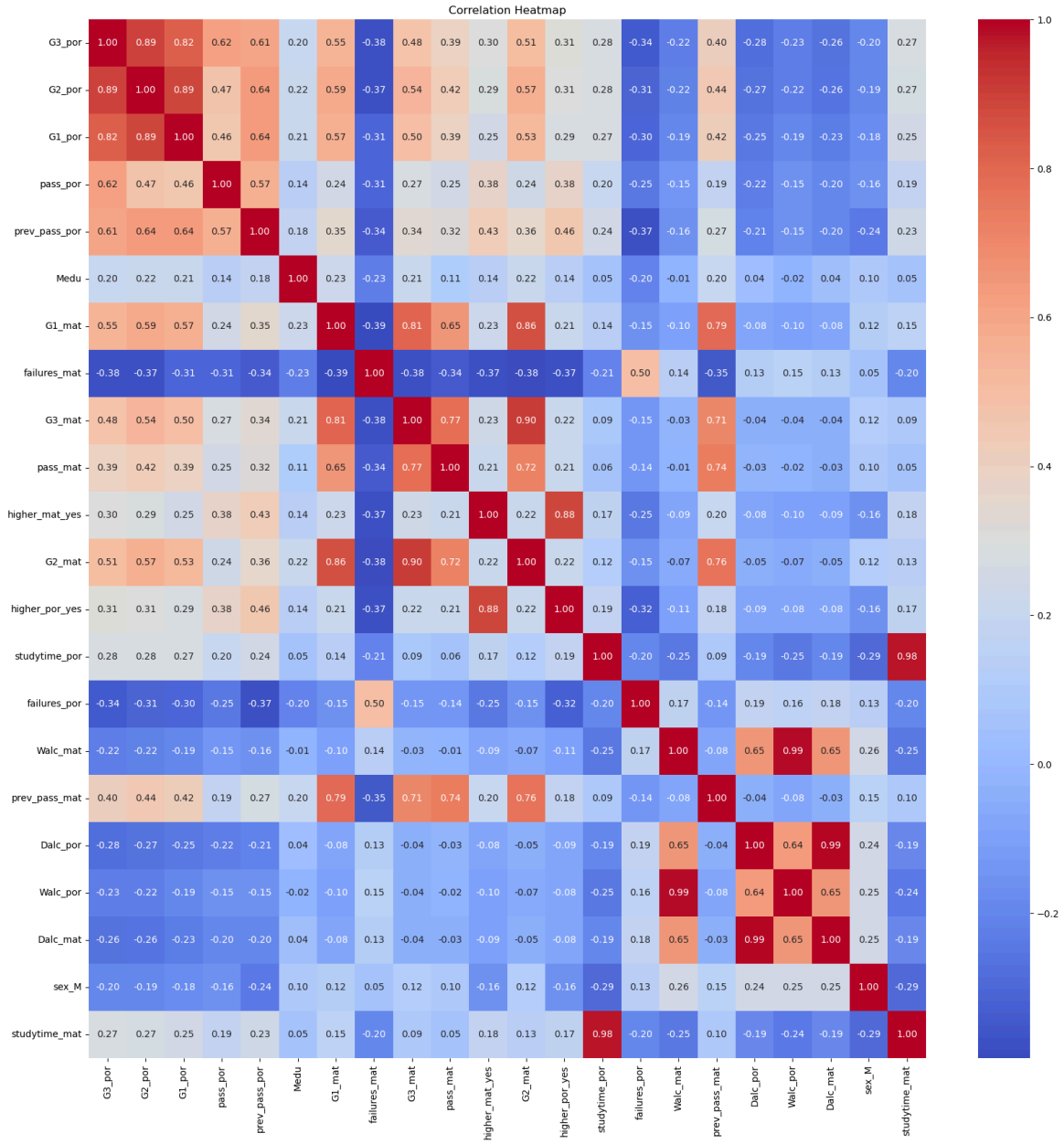
- 246 students passed both Math and Portuguese.
- 104 students failed Math but passed Portuguese.
- 9 students passed Math but failed Portuguese.
- 23 students failed both subjects.

These findings raise questions: Why is Math harder to pass? Why are there clusters of students with extremely low scores—especially zeros in the final term (G3)? Could this be linked to attendance, grading policies, or mid-year dropouts?

A correlation heatmap revealed key features associated with final grades:

- **Strong positive correlations**: Past performance (`G1`, `G2`), intention to pursue higher education (`higher_yes`), study time (`studytime`), and mother's education level (`Medu`).

- **Strong negative correlations**: Alcohol use during the week (`Dalc`) and on weekends (`Walc`), number of past failures (`failures`), and number of absences (`absences`).



Modeling

Pre-Processing

- All categorical variables were one-hot encoded (dropping one to avoid multicollinearity).
- Features were scaled for modeling.
- Dimensionality was reduced via PCA for exploratory analysis.
- The dataset was split into training and testing sets (70/30 split).

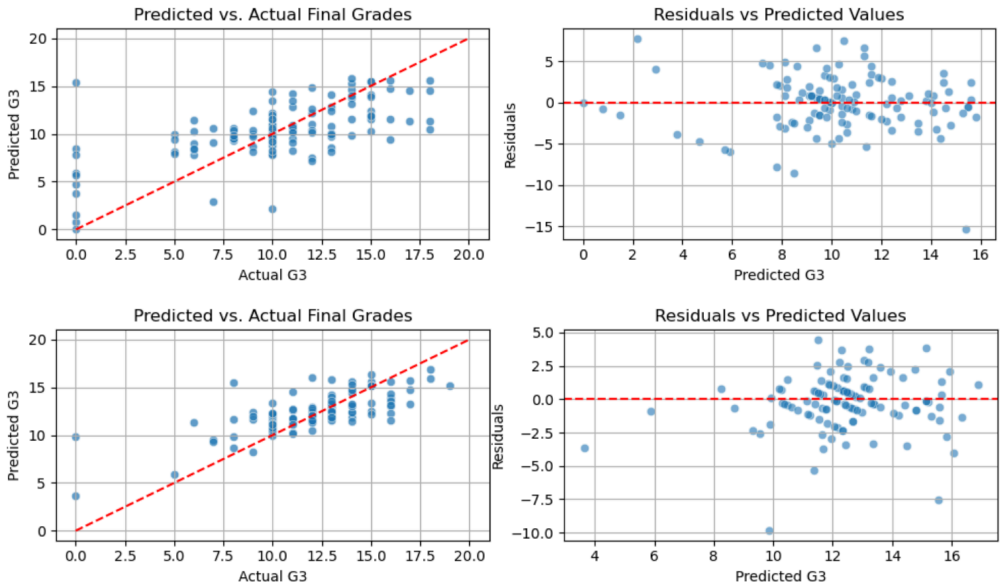
Regression Modeling

To predict final grades in Math and Portuguese, six regression models were tested:

- Linear Regression
- Ridge Regression
- Lasso Regression
- Gradient Boosting Regressor
- Random Forest Regressor
- XGBoost Regressor

Results:

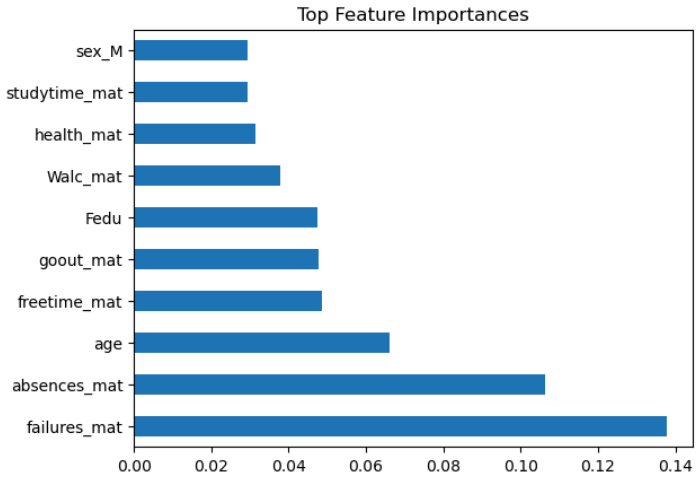
- For both subjects, **Gradient Boosting** had the best performance.
- **Math Grade (G3)**: RMSE = 4.11
- **Portuguese Grade (G3)**: RMSE = 4.22



Classification Modeling

To predict pass/fail outcomes, six classification models were used:

- Logistic Regression
- Random Forest Classifier
- Gradient Boosting Classifier
- XGBoost Classifier
- K-Nearest Neighbors
- Support Vector Classifier

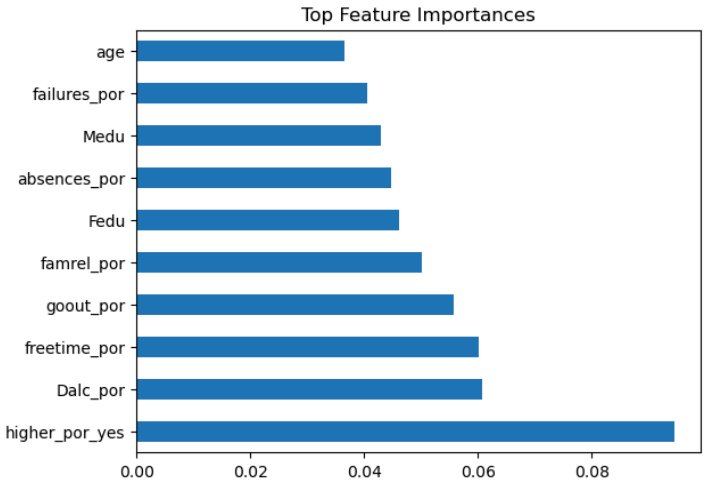


Math Classification

- Best model: **Random Forest**
- Mean cross-validated score: **0.745**
- Top features: `failures_mat`, `absence_mat`, `age`, `freetime_mat`, `goout_mat`

Portuguese Classification

- Best model: **Random Forest**
- F1 Score: **0.887**
- ROC AUC Score: **0.948**
- Top features: `higher_por_yes`, `Dalc_por`, `freetime_por`, `goout_por`, `famrel_por`



Conclusion

Takeaways

Key predictors of student success include:

- Past course failures
- Number of absences
- Age
- Higher education aspirations
- Weekday alcohol consumption
- Mother's education level
- Study time
- Family relationship quality
- Social habits (free time and going out)

Math appears to be more challenging for students, with more failures and sudden grade drops. This calls for further investigation into grading policies, curriculum rigor, and student engagement.

Recommendations

- **Early warning systems**: Flag students with prior failures or high absences for intervention.
- **Family engagement**: Improve home-school communication and offer family bonding opportunities.
- **Higher education support**: Create mentorship and scholarship programs to encourage college pathways.
- **Health education**: Raise awareness about alcohol's impact on academic performance.
- **Math-specific support**: Consider tutoring, curriculum review, and teacher training.

Further Research

- Qualitative interviews with students and teachers to understand sudden grade drops.
- Investigate school-level policy differences in grading and attendance enforcement.
- Expand the dataset to include more schools and longitudinal data across multiple years.
- Analyze the impact of mental health and external responsibilities (e.g., caregiving, work) on academic outcomes.